NIT-417

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of

Y. NONAKA et al

Serial No. 10/805,260 Group Art Unit: 2188

Filed: March 22, 2004 Examiner: M. Padmanabhan

For: STORAGE SYSTEM HAVING A PLURALITY OF INTERFACES

PETITION TO MAKE SPECIAL
UNDER 37 CFR §1.102(d) (MPEP §708.02(VIII))

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

June 3, 2005

Sir:

The Applicants petition the Commissioner to make the above-identified application special in accordance with 37 CFR §1.102(d).

In accordance with the requirements set forth in Manual of Patent Examining Procedure §708.02(VIII), the Applicants believe that all claims are directed to a single invention. If the Office determines that all claims are not directed to a single invention, Applicants will make an election without traverse as a prerequisite to the grant of special status.

Discussion of invention as claimed in the independent claims

The claimed invention, as set forth in independent claims 17 and 23, is generally directed to a storage system and method including both a storage area network (SAN) controller and network-attached storage (NAS) controller.

Claim 17 is directed to a storage system including a disk array and a controller which comprises a SAN controller and a NAS controller coupled to the SAN controller. After the SAN controller is initialized, the disk array is arranged to be initialized by a disk-array control program, and the SAN processor is arranged to transmit a completion notice of the initialization of the disk array to the NAS controller. Then, after reception of the completion notice of the initialization of the disk array, the NAS controller initializes the NAS memory and transmits a completion notice of the initialization of the NAS memory to the SAN controller. The SAN controller then transmits a command conversion program to the NAS memory by executing a NAS controller management program, and transmits a completion notice of the program transmission, whereupon the NAS controller starts execution of the command conversion program.

Claim 23 is directed to a method for operating a storage system, including steps of initializing the SAN controller,

initializing the disk array, transmitting a completion notice of the initialization of the disk array from the SAN controller to the NAS controller, initializing the NAS memory after reception of the completion notice of the initialization of the disk array, transmitting a completion notice of the initialization of the NAS memory to the SAN controller, transmitting a command conversion program stored in a disk in the disk array from the SAN controller to the NAS controller to be stored in the NAS controller, after reception of the completion notice of the initialization of the NAS memory by the SAN controller, by executing a NAS controller management program stored in the SAN memory, and transmitting a completion notice of the transmission of the command conversion program. Then, the NAS processor converts a file I/O command received from a NAS client computer into a block I/O command using the command conversion program stored in the NAS controller, transfers the block I/O command to the SAN controller, and the SAN controller accesses data stored in the plurality of disks according to the block I/O command. The SAN processor can also access data stored in the plurality of disks according to a block I/O command received from a SAN client computer.

Discussion of documents developed by pre-examination search

In satisfaction of the requirement for a pre-examination search as set forth in MPEP §708.02(VIII)(C), the Applicants state that a search has been made by the UK patent office with respect to claims in a corresponding UK patent application of similar scope to the claims in the present US application for which Special status is requested. The documents developed by the UK patent office search are listed below. The claimed subject matter is believed to be patentable over the teachings of these documents for the reasons set forth. One copy of each of these documents accompanies this Petition.

Published U.S. Patent Applications

2002/0083120	Soltis
2002/0161982	Riedel
2003/0225735	Weber

Other Publication

Daniel, Stephen. "Converging SAN and NAS Storage - A Comparison of Unified and Gateway Solutions." Network Appliance, Inc., White Paper, Oct. 2002.

Soltis, U.S. Patent Publication No. 2002/0083120 ("Soltis") discloses a shared storage distributed file system that is said to provide applications with transparent access to a storage area network (SAN) attached storage device. According to the patent publication (excerpted and paraphrased):

> This is accomplished by providing clients read access to the devices over the SAN and by requiring most write activity to be serialized through a network attached storage (NAS) server. Both the clients and the NAS server are connected to the SAN-attached device over the SAN. Direct read access to the SAN attached device is provided through a local file system on the client. Write access is provided through a remote file system on the client that utilizes the NAS server. A supplemental read path is provided through the NAS server for those circumstances where the local file system is unable to provide valid data reads.
>
> Consistency is maintained by comparing modification times in the local and remote file systems. Since writes occur over the remote file systems, the consistency mechanism is capable of flushing data caches in the remote file system, and invalidating metadata and real-

data caches in the local file system. It is possible to utilize unmodified local and remote file systems by layering over the local and remote file systems a new file system. The file system need only be installed at each client, allowing the NAS server file systems to operate unmodified. Alternatively, the new file system can be combined with the local file system.

The distributed file system utilizes aspects of a NAS server system along with a storage area network having at least one SAN-attached storage device. By combining these two architectures, the system has the benefits of fast data reads over a SAN as well as some of the consistency benefits of using a NAS server. The architectures are combined by creating separate data paths for write and read requests.

The write data-path is similar to the write data-path of prior art NAS, with the NAS storage device being replaced with a SAN-attached storage device accessed over a SAN. This is accomplished so that all write activities to the SAN attached storage device are serialized through one server, while still allowing each client write access to the volume stored on the SAN-attached device.

The primary read data-path is similar to the read data-path of prior art SAN environments, whereas the secondary read data-path is similar to the read data-path of prior art NAS environments. Since most reads pass directly from the SAN-attached storage device to the clients, the system takes full advantage of high-speed SAN protocols. In those instances where the primary read data path is not available, the system utilizes the secondary data path of typical NAS environments.

In a first embodiment, a new file system is loaded into each client. This file system is layered on top of separate local and remote file systems, which handle actual data transfers over the SAN and actual data transfers with the NAS server. No modification of the file systems of the NAS server is necessary.

In a second embodiment, the file system is merged with a local file system. In this embodiment, this combined file system is used on the client in conjunction with a remote file system that handles communication with the NAS server. The new, combined file system is also used on the NAS server as the local file system. By using this combined data system, this second embodiment

ensures that all clients and servers accessing the SAN-attached devices will be able to coexist.

In a third embodiment, the file system is merged with a local file system and given support to write directly to the SAN-attached devices. In this embodiment, the client coordinates with the server to maintain consistency while updating on-disk modes and block allocate tables. The client has multiple options concerning which data-path to transfer data; however, in a common scenario, the client transfers small files and small file requests across the LAN using NAS protocols and transfers large files and large file requests across the SAN using SAN protocols.

Soltis does not disclose or suggest that completion notices are transmitted and received between the SAN controller and NAS controller, as claimed in independent claims 17 and 23, such that after the SAN controller and the disk array are initialized, a command conversion program is prepared and transmitted to the NAS memory and executed.

Riedel, U.S. Patent Publication No. 2002/0161982 ("Riedel") discloses a system 200 that includes a NAS 210 and a SAN 220. According to the patent publication (excerpted and paraphrased):

The NAS 210 may be configured to provide access to data storage capabilities of the SAN 220 through a network 230. The network 230 may be configured to provide a communication channel between the NAS 210 and clients 240. The clients 240 may be implemented as personal computers, workstations, servers, and the like.

In one aspect, a NAS 220 may send optimization information 250 to the SAN 220. Subsequently, the optimization information may be utilized by the SAN 220 to optimize the blocks in the storage devices for performance reliability, etc. For example, a client 240 may execute a remove (or delete) command on a file (or directory) maintained by the network file system 215 on the NAS 210.

The NAS 210 may be configured to delete the file and update an available free block table with the freed blocks associated with the deleted file (or directory). The NAS 210 may be further configured to generate and transmit a freed block message, as an example of optimization information 250, listing the freed blocks from the deleted file to the SAN 220. As a result, the SAN 220 may be configured to update a current free block

table with the freed blocks in response to receiving the freed block message.

Subsequently, the SAN 220 may flush from the SAN 220, mark as unused by the SAN 220 or mark as allocated but unused any blocks listed in the current free block table. The sending of the optimization information, e.g., the freed block message, from the NAS to the SAN can be done in an "out of band" manner, without changing the native interface between the NAS and the SAN.

The optimization information do not affect the correctness of the data sent from the SAN to the NAS, only the performance of the responses from the SAN to the NAS.

Riedel also does not disclose or suggest that completion notices are transmitted and received between the SAN controller and NAS controller, as claimed in independent claims 17 and 23, such that after the SAN controller and the disk array are initialized, a command conversion program is prepared and transmitted to the NAS memory and executed.

Weber, U.S. Patent Publication No. 2003/0225735 ("Weber") discloses apparatus and method for providing transparent sharing of channel resources by multiple host machines

utilizing mixed mode block and file protocols. According to the patent publication (excerpted and paraphrased):

> A storage complex system implements I/O modules and controller elements including circuits capable of translating to permit common message passing for providing transparent mixed mode data transport. An object converting element, included in the system, is capable of converting between file-based message passing while implementing its file system on the block-based storage complex components, thereby allowing for transport of object based requests.
>
> A method for providing mixed mode data storage transport includes receiving the host commands and data requests for access. The received host command and data is interpreted, including translating the data into a common block storage message and determining if the host request is an object request. If the request is an object request, the request is routed to an object-converting element for conversion. Data forming the requested object is assembled into the object and subsequently transferred to the requesting I/O channel for transport to the requesting host device.

Weber, however, does not disclose or suggest that completion notices are transmitted and received between a SAN controller and a NAS controller, as claimed in independent claims 17 and 23, such that after the SAN controller and the disk array are initialized, a command conversion program is prepared and transmitted to the NAS memory and executed.

Daniel, "Converging SAN and NAS Storage - A Comparison of Unified and Gateway Solutions" (Daniel) discloses a "unified storage system" having a SAN unit that receives block I/O commands and a NAS unit that receives file I/O commands, wherein both file and LUN services are said to be built on top of a common storage virtualization layer. According to the White Paper:

> The lowest levels, disk, RAID, and buffer cache, are layered in the fashion of all storage systems. Above the buffer cache is the virtualization layer that manages space. Storage space is then made available to applications either as LUNs or as files.

Daniel is primarily directed to a software solution to problems of a prior gateway architecture. Daniel does not disclose or suggest that completion notices are transmitted and received between a SAN controller and a NAS controller, as claimed in independent claims 17 and 23, such that after the

SAN controller and the disk array are initialized, a command conversion program is prepared and transmitted to the NAS memor and executed.

Conclusion

The pre-examination search required by the MPEP "must be directed to the invention as claimed in the application for which special status is requested." MPEP §708.02 (VIII). The search performed in support of this Petition is believed to be reasonable and approved by the MPEP as noted above; however, the Applicants make no representation that the search covered every search area that may contain relevant prior art. Prior art of greater relevance to the claims may exist. The Applicants urge the Examiner to conduct his or her own complete search of the prior art, and to thoroughly examine this application in view of the prior art cited above and any other prior art that may be located in the Examiner's independent search.

Further, while the Applicants have identified certain portions of each cited reference in order to satisfy the requirement for a "detailed discussion of the references, which discussion points out, with the particularly required by 37 C.F.R. §1.111(b) and (c), how the claimed subject matter is patentable over the references" (MPEP §708.02(VIII)), the Examiner should not limit review of these documents to the identified portions, but rather is urged to review and consider the entirety of each reference.

In conclusion, the Applicants submit that the foregoing discussion demonstrates the patentability of the claimed invention over the closest known prior art. Accordingly, the requirements of 37 CFR §1.102(d) having been satisfied, the Applicants request that this Petition be granted and that the application be examined according to prescribed procedures.

A Credit Card Payment Form in the amount of $130.00 is submitted herewith, in satisfaction of the fee set forth in 37 CFR §1.17(h). The Commissioner is hereby authorized to charge any additional payment due, or to credit any overpayment, to Deposit Account No. 50-1417.

Respectfully submitted,

Daniel J. Stanger
Registration No. 32,846
Attorney for Applicants

MATTINGLY, STANGER, MALUR & BRUNDIDGE, P.C.
1800 Diagonal Rd., Suite 370
Alexandria, Virginia 22314
(703) 684-1120
Date: June 3, 2005

NetworkAppliance®
The evolution of storage.™

Converging SAN and NAS Storage

A Comparison of Unified and Gateway Solutions

Stephen Daniel | October 2002 | TR 3199

WHITE PAPER

Network Appliance™ technology and expertise solve a wide range of data storage challenges for organizations, adding business value and enabling them to create and sustain a competitive advantage.

Abstract

Market research shows that network-attached storage (NAS) plays an ever-increasing role in storage of enterprise mission-critical data. Vendors of storage systems are responding by offering an ever-widening assortment of NAS storage systems. Over time, the industry will respond by offering systems that unify storage area network (SAN) storage with NAS storage.

First-generation solutions converge SAN and NAS storage by using a NAS device as a gateway to a SAN-based storage system. Network Appliance is the first vendor to unify SAN and NAS in a fabric-attached storage (FAS) system with both classes of protocols integrated as peers in a single, unified architecture. This paper compares gateway architectures with unified architectures and discusses the relative advantages of each approach.

Table of Contents

Introduction

The dramatic rise in the use of network-attached storage (NAS) in the enterprise challenges each vendor of enterprise-class storage systems to provide both NAS and SAN technologies. Vendors of SAN storage often initially respond to this market pressure by offering distinct NAS and SAN product lines. Over time, they seek to converge those products into a single product line. Eventually, these converged products will begin to unify. Each storage system will have to operate in both the SAN and NAS worlds.

Network Appliance implements unified SAN and NAS by building both styles of protocols into a single, integrated storage solution. This paper compares Network Appliance's unified approach with the more common approach of using a gateway device to form a bridge between SAN and NAS protocols.

This paper will show how and why a unified approach yields benefits in investment protection, total cost of ownership (TCO), performance, and scaling.

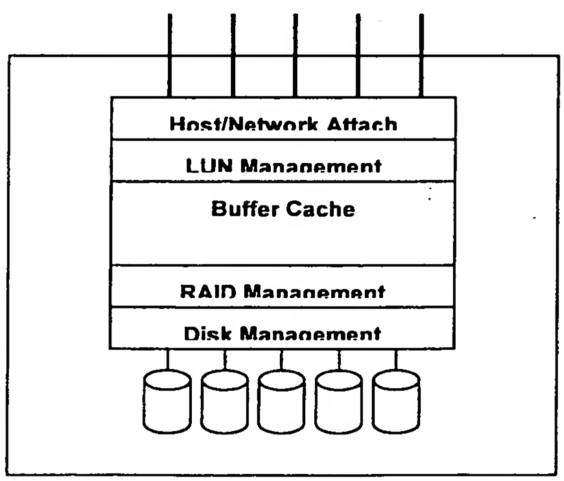
Background

SAN Technologies

All storage systems combine numerous disks into arrays for availability and performance. Typically, a collection of disks is grouped into a RAID array to protect the data. A buffer cache provides additional performance. SAN and NAS systems differ in how this storage is presented to the host systems.

SAN systems carve the storage into logical units (LUN) that are presented to the host in much the same manner as a single disk would be presented. LUNs are relatively few in number, relatively large, typically fixed in size, and accessible only in whole blocks.

This diagram illustrates these layers, from disks up to host attach.



Architecture of a typical SAN system

Because the SAN system is built to service block-oriented protocols, it is free to assume their constraints. SAN systems provide for manual creation and management of LUNs. The blocks of LUNs

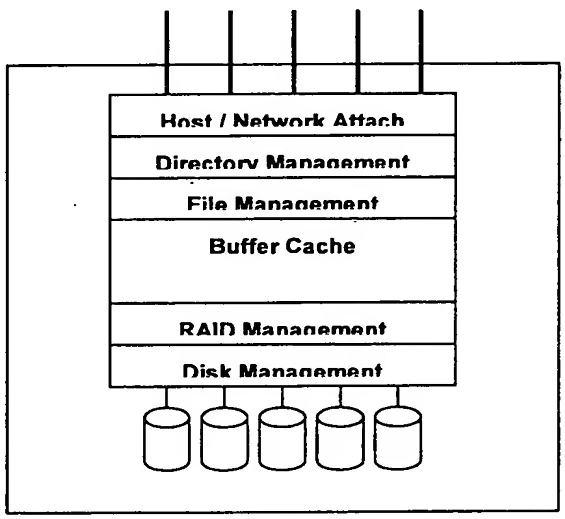
are statically mapped onto the blocks of the underlying disk subsystem. Achieving high performance in this environment is relatively straightforward and dependent mostly on the performance of the hardware used.

NAS Technologies

NAS storage systems are built using the same basic components as SAN systems. Collections of disks are built-in to RAID arrays and the resulting storage is made available to various hosts.

NAS systems differ sharply from SAN systems by presenting the storage as a dynamic set of files, rather than a static set of LUNs. Files can easily be created, destroyed, or grown. Files live in hierarchical directories. NAS devices manage thousands or millions of files, rather than tens or hundreds of LUNs. For this reason, NAS systems are optimized to provide sophisticated space management and performance under dynamic workloads. The software required to implement a full-featured file system and achieve good performance with a file-oriented NAS protocol is much more complex than the software required for SAN systems.

This diagram shows the layers of a NAS system, where complex file- and directory-management services replace the straightforward LUN management of a SAN system.



Architecture of a typical NAS system

Converging SAN and NAS Technologies

Collocation

The simplest approach to surviving in a world of both SAN and NAS protocols is to collocate SAN and NAS devices in the corporate IT infrastructure. This type of collocation has been a fact of the corporate data center for many years. This approach works well and system designers can optimize each class of storage system to provide a single class of service. Such an implementation provides the IT manager with flexibility in vendors and best-of-breed SAN and NAS implementations, but there are substantial drawbacks.

The chief drawback of separate SAN and NAS implementations is the lack of any convergence. An IT manager with both classes of systems may face all of these problems:

Separate devices will have distinct management interfaces.

They may have different and incompatible backup strategies.

Disk capacity cannot be shared across these platforms.

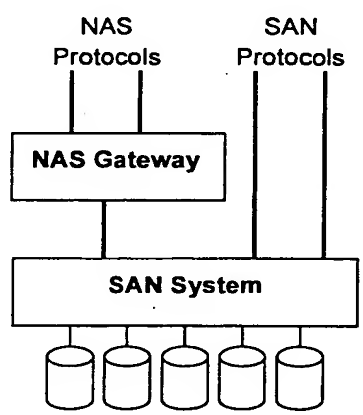
Since each disk is located in either a SAN or a NAS device, application workloads cannot be load-balanced across the available disk throughput.

Migrating an application from one environment to another may require substantial infrastructure changes.

All of these factors contribute to a higher cost of ownership and to higher administrative labor costs for these systems.

Converged NAS and SAN Systems

First-generation systems offer convergence of SAN and NAS by providing a NAS gateway in front of a SAN device. If such a system is implemented well, it offers advantages over simply collocating a NAS and a SAN. The following figure shows a sample architecture for a converged system built with a gateway.



NAS Gateway Architecture

Gateway technologies offer several advantages over simple collocation of distinct systems.

A gateway system can allow storage to be reallocated from SAN to NAS or NAS to SAN.

A well-implemented gateway system can make the throughput of all disks available to either style of application at a point in time when one class of application has a burst of load.

Time-to-market. A storage vendor may purchase a file system from an operating system vendor, package the software into a box, and sell the result as a NAS system or a NAS/SAN gateway.

The alternative, developing or porting a sophisticated file system into a SAN solution, can consume years of development effort.

A gateway approach offers the possibility of functional multiprocessing. One CPU is dedicated to running the file system, while another provides SAN and RAID services. If carefully implemented, this approach can make effective use of more CPU power than would otherwise be available.

A gateway may be an economical way to enhance the value of an existing, installed storage system. For example a large SAN system can begin servicing NAS applications through the use of a gateway.

Unfortunately, the disadvantages of a gateway approach are numerous. Some of the issues are fundamental to the architecture, and some are imposed by the implementations that are commonly used.

Placing a NAS gateway in front of a SAN system will create a number of problems:

The NAS gateway and the SAN system will typically have distinct management interfaces. Since the most common implementation involves purchasing a file system, full integration of the acquired management software into the SAN management framework often takes years. For this reason a NAS/SAN gateway usually fails to provide much relief from the higher costs associated with separate SAN and NAS solutions.

NAS applications will operate at the slower of the two systems' speeds, wasting processing capability on the other. Often the NAS gateway will be the bottleneck.

Split and redundant caches. Both SAN and NAS devices require substantial RAM caches to operate at peak efficiency. When NAS is implemented as a gateway in front of a SAN system, each must manage its own cache, often caching identical data.

Extra data movement. Moving data from disk to the SAN system, then to the NAS gateway, and finally to a NAS application is fundamentally slower than the data movement in a system that connects the disks directly to the NAS system.

Inability to share hardware resources. For example, iSCSI and NFS can both make use of an Ethernet interface, while a gateway product will typically be unable to serve both protocols over the same interface. NAS requests will be directed to the NAS gateway, while iSCSI will be directed to a distinct interface on the SAN system.

In addition to these architectural issues most, if not all, current NAS over SAN gateway products suffer from the following implementation limitations:

Storage partitioning. Frequently the NAS partitions are segregated to a distinct set of disks. During periods of peak NAS activity, the system's performance will be limited to the performance of that subset of the spindles. During periods of peak SAN activity, the NAS disks do not contribute to performance.

Poor integration of file system layout and RAID services. File systems are designed to optimize performance. Their optimizations are based on a set of assumptions about the underlying disk technology. Rarely are these assumptions well matched to the underlying RAID system. Differences in fundamental block size, block alignment, and support for sequential and random

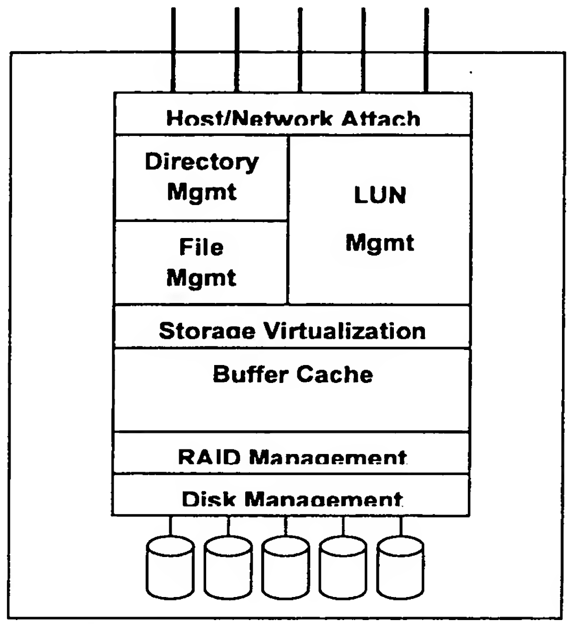
access often limit the effective performance of the RAID subsystem when accessed by the NAS gateway.

Code bloat. A NAS gateway product must solve a complex problem. However, such gateways are often *more* complex than they need to be. Because file systems are large and expensive software products, nearly all NAS gateways are built using file system technology developed for some other purpose. Some vendors use a version of Linux® or Windows NT® to create their NAS device. The use of a full-featured operating system, even if most of the features are hidden, carries with it numerous inefficiencies. Building the NAS system out of a full-function operating system also creates the risk of buying the security problems of the underlying operating system, in addition to its file system technology.

Unified SAN and NAS Technologies

As an alternative to gateways, Network Appliance approaches converged storage through its pioneering unified storage systems. Rather than implementing a gateway between NAS and SAN, Network Appliance filers build both file and LUN services on top of a common storage virtualization layer.

The following diagram shows the system's software architecture. The lowest levels, disk, RAID, and buffer cache, are layered in the fashion of all storage systems. Above the buffer cache is the virtualization layer that manages space. Storage space is then made available to applications either as LUNs or as files.



Network Appliance's Unified Storage Architecture

The benefits of this approach are substantial.

Unified cache. A single cache is used to cache file data for the NAS protocols, block data for the SAN protocols, and to optimize RAID I/O services.

Minimal data movement. Both block protocols and RDMA-capable NAS protocols are implemented with zero copies. Data is moved by DMA between the host and cache memory, and by DMA between cache memory and the disks. No copies are required, and only the minimum of a single hop. This technology maximizes throughput and minimizes latency.

RAID-optimized virtualization layer. Both file services and LUN services are built on top of the same virtualization layer. Network Appliance's patented WAFL® technology optimizes both block and file I/O operations for the underlying integrated RAID layer. This technology provides the performance superior to traditional RAID1/0 technologies with the space efficiencies of RAID5 arrays.

Symmetric multiprocessing. Network Appliance's multi-CPU products can use any CPU for any purpose, preventing the kind of bottlenecks that come from functional multiprocessing.

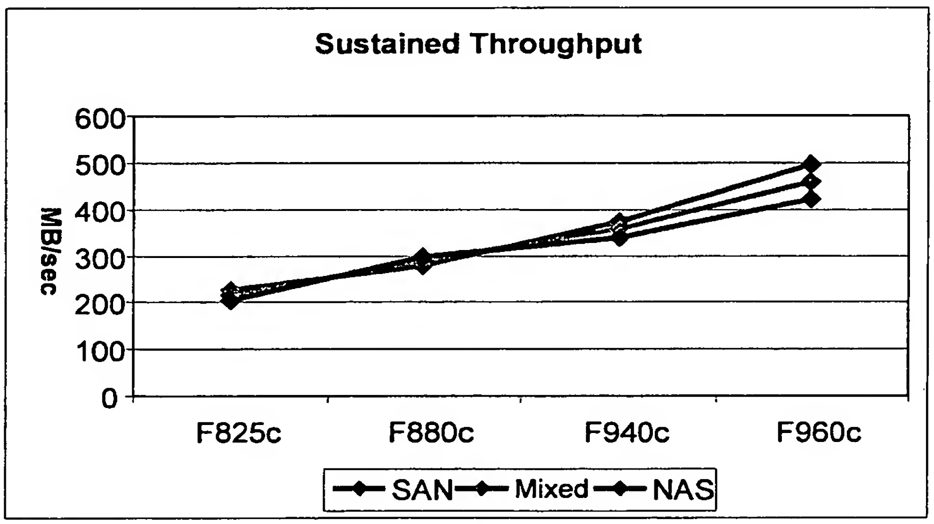
Integrated storage management. Network Appliance storage systems combine large numbers of disks into volumes. The space in these volumes can be made available as LUNs or files or a mixture. These volumes provide a wealth of features to the LUNs or files that live within them.

Any LUN created in a Network Appliance volume is automatically load balanced across all of the underlying disks.

All LUNs and files can be protected by Network Appliance's Snapshot™, SnapRestore®, and SnapMirror® technologies.

LUNs can be made available as files, and files can be converted into LUNs. These in-place operations allow administrators to use the most convenient protocol to access and manage their data.

The performance graph below demonstrates the flexibility inherent in Network Appliance's integration strategy. This graph compares throughput for four different models of Network Appliance filers using SAN, NAS, and mixed workloads.



A Comparison of SAN and NAS Throughput for Several Network Appliance Filers

As clearly shown by the graph, the performance of SAN and NAS protocols are comparable. The different models contain different optimizations that may slightly favor one protocol family over another, but neither protocol family shows a distinct advantage over the entire product family. Further, as shown by the mixed workload performance, available throughput can be efficiently shared among protocols.

Conclusions

Network Appliance storage subsystems provide industry-leading convergence of SAN and NAS into fabric-attached storage devices. These devices provide excellent services in pure SAN, pure NAS, and mixed environments. Clearly differentiated from first-generation converged systems built from

gateways, Network Appliance's approach offers integrated management, lower cost of ownership, and superior performance.